AWARD NUMBER: W81XWH-11-1-0713

TITLE: Identification of Novel Inherited Genetic Markers For Aggressive PCa in European and African Americans Using Whole Genome Sequencing

PRINCIPAL INVESTIGATOR: Jielin Sun

CONTRACTING ORGANIZATION: Wake Forest University
Winston Salem, NC 27157

REPORT DATE: October 2014

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE October 2014	2. REPORT TYPE Final	3. DATES COVERED 22 Aug 2011 – 21 Aug 2014

4. TITLE AND SUBTITLE
Identification of Novel Inherited Genetic Markers for Aggressive PCa in European and African Americans Using Whole Genome Sequencing

5a. CONTRACT NUMBER

5b. GRANT NUMBER W81XWH-11-1-0713

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jielin Sun, PhD; Siqun Lilly Zheng, MD

E-Mail: jisun@wakehealth.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wake Forest University

Winston Salem, NC 27157

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Prostate cancer (PCa) is the most common cancer and the second leading cause of cancer death among men in the US. While most PCa patients have an indolent form of the disease that may not even require treatment, about 10-15% of PCa patients have an aggressive form that may progress to metastases and death, thus requiring intensive treatment. Several clinical variables such as PSA levels, Gleason grade and TNM stage are good predictors for disease with poor clinical outcomes; however, their predictive performance needs to be improved. Our inability to reliably distinguish between these two forms of PCa, early on in the course of the disease has resulted in the over-treatment of many and under treatment of some. The identification of additional markers, including genetic variants will improve our ability to distinguish aggressive from indolent forms of PCa and to better understand the racial disparity of PCa that exists between Europen Americans (EA) and African Americans (AA). In this DOD proposal, we hypothesized that multiple rare sequence variants in the genome may increase aggressive PCa risk. Through a GWAS of rare variants based on three existing populations from Johns Hopkins Hospital, Michigan and CAPS population from Sweden, with a total of 1,919 PCa cases, including 470 aggressive PCa cases and 1,449 indolent PCa cases using Illumina's Human Exome BeadChip, we have identified and confirmed several novel rare variants in the *INPP5D* gene and *HINFP* gene that are associated with aggressive PCa in both Caucasians and African American men. The newly identified variants can provide more insight into the etiology of aggressive PCa and provide potential effective targets for therapy of aggressive PCa.

15. SUBJECT TERMS
Prostate cancer, indolent, lethal, aggressive, sequence variants

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	23	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

INTRODUCTION

While most prostate cancer (PCa) patients have an indolent form of the disease that may not even require treatment, about 10-15% of PCa patients have an aggressive form that may progress to metastases and death, thus requiring intensive treatment. Several clinical variables such as PSA levels, Gleason grade, and TNM stage are good predictors for disease with poor clinical outcomes; however, their predictive performance needs to be improved. Our inability to reliably distinguish between these two forms of PCa, early on in the course of the disease has resulted in the over-treatment of many and under treatment of some. Another dilemma is a large difference in PCa risk, especially aggressive PCa, between races. African Americans (AAs) have the world's highest incidence of PCa and are twice as likely, as compared with Caucasians to die of the disease. Inherited markers of aggressive PCa could be used for screening and diagnosis of aggressive PCa at an early stage while reducing over-diagnosis and treatment for others. The overall hypothesis is that inherited sequence variants in the genome are associated with a lethal (aggressive) form of PCa but not indolent PCa, and the difference in these variants between races may contribute to higher incidence of and mortality from aggressive PCa in AA.

In this DOD proposal, we proposed: 1) To discover novel inherited genetic variants in the genome that may be associated with aggressive but not indolent PCa using an exome array approach; **2)** To confirm the novel genetic variants using mass spectrometry directed sequencing; and **3)** To perform association tests of implicated genetic variants among 1,500 most aggressive PCa and 1,500 least aggressive (i.e. indolent) PCa.

BODY

Approved Revised Statement of Work:

Aim 1. To discover novel inherited genetic variants in the genome that may be associated with aggressive but not indolent PCa using a WGS approach.

Step by Step method and expected results

1. **Months 1-6:** Preparation of the study, including regulatory review, IRB approval and other logistical issues
2. **Months 7-12**: Perform exome SNP array analysis for 400 (200 aggressive PCa and 200 indolent PCa) cases in EAs and 400 (200 aggressive PCa and 200 indolent PCa) cases in AAs from Johns Hopkins Hospital.
3. **Months 13-18:** Perform exome SNP array analysis for 200 (100 aggressive PCa and 100 indolent PCa) cases in EAs and 200 (100 aggressive PCa and 100 indolent PCa) cases in AAs from Johns Hopkins Hospital. Perform statistical and bioinformatics analysis for the combined dataset of 600 aggressive PCa cases and 600 indolent PCa cases.

Outcome and deliverables

We expect to identify a certain number of novel rare variants most likely associated with aggressive but not indolent PCa.

Aim 2. To confirm the genetic variants implicated in Aim 1 using Sequenom

Step by Step method and expected results

1. **Months 19-22:** Genotyping the top rare mutations among the additional PCa samples using Sequenom
2. **Months 23-24**: Confirmation analysis of the top SNPs

Outcome and deliverable
We expect that a subset of the top rare mutations will be confirmed using the Sequenom platform.

Aim 3. To perform association tests of selected genetic variants among 1,500 most aggressive PCa and 1,500 most indolent PCa.

Step by Step method and expected results

1. **Months 25-26:** Genotyping ~100 SNPs in 1,500 most aggressive PCa and 1,500 most indolent PCa patients
2. **Months 27-28:** Perform association test of these SNPs with aggressiveness of PCa using a logistic regression model
3. **Months 29-36:** Final analysis and preparation of papers

Outcome and deliverable
We expect to identify several novel rare mutations that are associated with aggressiveness of PCa using exome SNP array analysis. We will prepare and submit papers reporting the major results from the study.

Detailed report

Study design modification. In our initial report for year 3, we proposed to perform association tests of selected ~100 top genetic variants among additional aggressive PCa and indolent PCa from JHH population using Sequenom Platform. During year 3, we were able to obtain access to the Exome BeadChip array data for two additional Caucasian populations (Michigan and CAPS) with 328 additional aggressive PCa cases and 814 indolent PCa cases. Therefore, we also conducted a genome-wide association analysis for rare variants with PCa aggressiveness in those two populations. We also conducted a pooled analysis using all three populations with Exome array data with a total of total of 1,919 PCa cases, including 470 aggressive PCa cases and 1,449 indolent PCa cases. Only rare variants that were implicated in all three populations were followed up for additional confirmation in additional 2,355 PCa cases with 1,076 aggressive PCa and 1,291 indolent PCa cases from CAPS population. Compared with our original study design, the new design greatly improved

our statistical power due to increased sample sizes. We were also able to decrease the number of false positive results by including two more populations to compare the association results for all the rare variants on the Exome Array chip.

Study Subjects. Subjects included in the John Hopkins (JHH) study were recruited during Jan. 1999 to Dec. 2008. All of them underwent radical prostatectomy for treatment of prostate cancer. Details of this study have been described in previous publications. In this study, aggressive prostate cancer was defined as: 1) Gleason Score ≥8; or 2) Gleason Score =7, with the most prevalent pattern being 4; or 3) stage T3b or higher; or 4) involvement of regional lymph nodes; or 5) presence of distant metastasis. Otherwise, the cancers were classified as non-aggressive prostate cancer. In this study, a total of 1,177 subjects (including 777 EA and 400 AA samples) from JHH study were genotyped using the Illumina Human Exome BeadChip platform. In addition, 772 PCa cases of AA descent including 388 subjects with aggressive PCa and 384 indolent PCa cases were genotyped to replicate the rare variants that were implicated in the first stage of Exome Array analysis based on AA population.

The second population included subjects recruited in Sweden from the CAPS study, which were diagnosed from Jul. 2001 and Oct. 2003. Details of this study have been described in previous publications. In the CAPS study, aggressive prostate cancers were defined as: 1) Gleason Score ≥8; or 2) stage T3 or higher; or 3) involvement of regional lymph nodes; or 4) presence of distant metastasis; or 5) serum PSA >50 ng/ml. Otherwise, the cancers were classified as non-aggressive prostate cancer. In this study, 446 subjects from CAPS study were genotyped by ExomeArray. Among them, 149 subjects were aggressive prostate cancer patients while 297 patients had anon-aggressive form. In addition, 2,355 cases with 1,064 aggressive PCa cases and 1,291 indolent PCa cases were genotyped to replicate the rare variants that were implicated in all the three populations.

The third study population included subjects recruited by the University of Michigan. The definition of prostate cancer aggressiveness in the Michigan population is exactly the same as in the JHH study. In this study, 864 subjects from Michigan study were genotyped using the Human Exome BeadChip platform. Among them, 179 subjects were aggressive prostate cancer patients while 517 subjects had a non-aggressive form of the disease.

Genotyping and Quality Control. Genotyping of samples in the first stage was conducted using the Illumina Human Exome BeadChip at the Center for Cancer Genomics, Wake Forest University School of Medicine. A total of 247,870 genetic variants were included in the ExomeArray. Those polymorphic SNPs were used for sex and IBS check of all subjects using PLINK software (Purcell 2007). In addition, polymorphic SNPs were also used to estimate the missing rate per individual. In each stage, subjects with genotyping missing rate >5% were removed from further analysis. For subjects in stage 1 with exome data available, IBS check and sex check were also performed. SNPs with missing rates >2% in subjects passed quality control (QC) were removed from further analysis.

Top variants selected to be confirmed in CAPS and JHH were genotyped using

TheSequenom MassArray system. The MassARRAY system is composed of four parts: assay components for primer extension reactions, SpectroCHIP arrays (silicon chips consisting of 384 elements on which extension products are spotted), a MALDI-TOF mass spectrometer, and SpectroTYPER software for automated scoring of SNP alleles. SNPs were genotyped by multiplexing of ~15 SNPs that were genotyped in a single analysis. PCR were performed using standard conditions. Reactions were scaled down to 5ul and used less than 10ng of genomic DNA per multiplexing group. All of the post-PCR reactions were performed using proprietary MassARRAY components and equipment for increased consistency and accuracy.

Several measures were taken to ensure the quality of the genotype data using the Sequenom platform. All samples were aliquoted into genotyping plates and each was assigned a unique barcode. All plates contained an asymmetric distribution of control samples (CEPH and water) throughout the plate in order to prevent plate flipping and to allow for unique identification. To ensure high quality genotyping, 2 quality control samples and 2 blank samples were included in each 96-well plate. Cases and controls were randomly included in each plate and their status was blinded to laboratory personnel.

Bioinformatics analysis (Variant effect prediction): All coding nonsynonymous variants were assessed for potential effect by Polymorphism Phenotyping version 2 (PolyPhen2), which is a tool for predicting the possible impact of an amino acid substitution on the structure and function of a human protein. For a given variant, PolyPhen2 calculates a Naïve Bayes posterior probability that the mutation is damaging and then appraised qualitatively as benign, possibly damaging, or probably damaging (Adzhubei 2010).

Statistical analysis for single SNP effect. Principal components analysis was conducted to detect potential population stratification by EIGENSTRAT software (Price 2006). The top five eigenvectors which indicated ancestral heterogeneity within a group of individuals were adjusted as covariates in the multivariate logistic regression analysis.

All polymorphic genetic variants that passed QC were evaluated for associations with prostate cancer aggressiveness. For genetic variants with any of the genotype counts ≤5, Fisher's exact test was applied to investigate potential association. For genetic variants with genotype counts >5, multivariate logistic regression analysis was conducted assuming an additive genetic model, adjusting for age-at-diagnosis and the top five eigenvectors. All analyses were performed using the PLINK software package (Purcell 2007).

Gene-based analysis: We used a novel statistical approach called Sequence Kernel Association Test (SKAT), to conduct gene-based analysis of rare variants for aggressive PCa. SKAT is a supervised and flexible regression method to test for association between rare variants in a gene or genetic region and a continuous or dichotomous trait. Compared to other methods of estimating the joint effect of a subset of SNPs, SKAT is able to deal with variants that have different direction and magnitude of effects, and allows for covariate adjustment (Wu 2011). In addition,

SKAT can also avoid arbitrary selection of threshold in burden test. Moreover, SKAT is computationally efficient, compared to a permutation test, making it feasible to analyze the large dataset in our study.

Results

EA population. Detailed clinical and demographic characteristics for the study population in stage 1 were presented in Table 1.

Table 1. Clinical and Demographic Characteristics of Subjects in Stage 1.

	JHH # (%)		MI # (%)		CAPS # (%)	
Characteristics	Agg (N=142)	Non-Agg (N=635)	Agg (N=179)	Non-Agg (N=517)	Agg (N=149)	Non-Agg (N=297)
Age at enrollment (Year)						
Mean (sd)	51.5 (3.9)	49.29 (4.44)	NA	NA	NA	NA
Age at disgnosis						
≤ 55	NA	NA	178 (99.4)	517 (100)	48 (32.2)	93 (31.3)
> 55	NA	NA	1 (0.6)	0	101 (67.8)	204 (68.7)
Missing	NA	NA	0	0	0	0
Family History (first-degree relatives)						
No	125 (88.0)	551(86.8)	NA	NA	105 (70.5)	184 (62.0)
Yes	15 (10.6)	66 (10.4)	NA	NA	41 (27.5)	109 (36.7)
Missing	2 (1.4)	18 (2.8)	NA	NA	3 (2.0)	4 (1.3)
PSA levels at diagnosis for cases or at enrollment for controls (ng/ml)						
≤ 4	21 (14.8)	224 (35.3)	12 (6.7)	164 (31.7)	7 (4.7)	60 (20.2)
4.01-9.99	78 (54.9)	357 (56.2)	89 (49.7)	281 (54.4)	25 (16.8)	157 (52.9)
10-19.99	23 (16.2)	45 (7.1)	30 (16.8)	39 (7.5)	22 (14.8)	55 (18.5)
20-49.99	18 (12.7)	4 (0.6)	20 (11.2)	4 (0.8)	25 (16.8)	23 (7.7)
50-99.99	0	0	19 (10.6)	1 (0.2)	25 (16.8)	0
≥100	0	0	0	0	43 (28.9)	0
Missing	2 (1.4)	5 (0.8)	9 (5.0)	28 (5.4)	2 (1.3)	2 (0.7)
T-stage						
T1	0	0	0	1 (0.2)	20 (13.4)	173 (58.2)
T2	47 (33.1)	512 (80.6)	71 (39.7)	467 (90.3)	26 (17.4)	122 (41.1)
T3a	53 (37.3)	123 (19.4)	33 (18.4)	49 (9.5)	0	0
T3b	41 (28.9)	0	33 (18.4)	0	0	0
T3c	0	0	0	0	0	0
T3x	1 (0.7)	0	0	0	83 (55.7)	0
T4	0	0	3	0	18 (12.1)	0
TX	0	0	0	0	0	0
Missing	0	0	39 (21.8)	0	2 (1.3)	2 (0.7)
N-stage						
N0	119 (83.8)	627 (98.7)	119 (66.5)	410 (79.3)	36 (24.2)	60 (20.2)
N1	16 (11.3)	0	26 (14.5)	0	22 (14.8)	0
NX	1 (0.7)	8 (1.3)	20 (11.2)	107 (20.7)	91 (61.1)	237 (79.8)

Missing	0	0	14 (7.8)	0	0	0
M-stage						
M0	0	0	81 (45.3)	257 (49.7)	76 (51.0)	110 (37.0)
M1	0	0	15 (8.4)	0	45 (30.2)	0
MX	142 (100)	635 (100)	72 (40.2)	260 (50.3)	28 (18.8)	187 (63.0)
Missing	0	0	11 (6.1)	0	0	0
Gleason (biopsy)						
≤ 4	0	0	0	6 (1.2)	0	21 (6.7)
5	0	8 (1.3)	0	21 (4.1)	9 (6.0)	49 (16.5)
6	1 (0.7)	420 (66.1)	6 (3.4)	272 (52.6)	25 (16.8)	163 (54.9)
7 (3+4)	16 (11.3)	207 (32.6)	16 (8.9)	218 (42.2)	0	60 (20.2)
7 (4+3)	75 (52.8)	0	84 (46.9)	0	48 (32.2)	0
7 (total)	91 (64.1)	207 (32.9)	100 (55.9)	218 (42.2)	48 (32.2)	60 (20.2)
8	31 (21.8)	0	31 (17.3)	0	22 (14.8)	0
9	19 (13.4)	0	35 (19.6)	0	31 (20.8)	0
10	0	0	3 (1.7)	0	3 (2.0)	0
Missing	0	0	0	0	11 (7.4)	4 (1.3)

In stage 1 of the Exome Array analysis, a total of 247,870 genetic variants were included in this ExomeArray. Among them, 92,173, 88,087 and 71,435 genetic variants were polymorphic in JHH, Michigan and CAPS populations, respectively. For polymorphic genetic variants, only those with a missing rate >0.98 in subjects passed QC were kept for further statistical analyses, including 91,998 variants in JHH, 87,879 variants in MI and 71,220 variants in CAPS. 79,729, 60,243, 57,126 genetic variants had an MAF < 0.1 in the JHH, Michigan and CAPS population, respectively.

Association Analysis for single variants.

We did not observe any association between genetic variants and PCa aggressiveness achieved genome-wide significance (P<5E-7) in JHH, Michigan or CAPS populations. In the JHH population, 47 variants were significantly associated with PCa aggressiveness with p-value < 1E-3, including 13 rare variants with minor allele frequency (MAF) < 0.05, and 34 common ones (MAF ≥ 0.05). In the Michigan population, we found 27 variants significantly associated with PCa aggressiveness (p-value < 1E-3), including 11 rare ones and 16 common ones. In the CAPS population, we identified 18 variants significantly associated with PCa aggressiveness (p-value < 1E-3), including 7 rare ones and 11 common ones. No variants are significantly associated with PCa aggressiveness with p-value < 1E-3 in all three populations.

We therefore conducted association analysis in a pooled sample of all three populations. A total of 39 genetic variants achieved P-value<0.001 in the pooled analysis. Among them, 31 variants had effects in the same direction in all three populations. We further examined the 31 variants and found 11 of them 11 variants were showed significant association (P < 0.05) in at least 2 populations. The association results of the 11 variants in each population and the pooled analysis were presented in Table 2.

Table 2. Associations between single variants and prostate cancer aggressiveness in JHH, Michigan and CAPS populations in Exome Array analysis

Population	SNP	CHR	BP	Gene	A1/A2	MAF Agg	MAF indolent	Genotype Agg	Genotype Indolent	OR	P-value
JHH											
	bs1_156907115	1	156,907,115	ARHGEF11	G/A	0.408	0.346	17/77/42	80/272/273	1.38	2.30E-02
	bs2_233990575	2	233,990,575	INPP5D	A/G	0.007	0.000	0/2/134	0/0/625		3.18E-02
	rs115393439	2	233,990,592	INPP5D	C/A	0.018	0.003	0/5/131	0/4/621	5.83	1.19E-02
	rs464494	5	76,003,258	IQGAP2	A/G	0.375	0.457	22/58/56	120/331/174	0.70	1.42E-02
	rs61740965	5	81,608,563	ATP6AP1L	G/A	0.022	0.004	0/6/130	0/5/620	5.62	6.40E-03
	rs10274334	7	47,925,331	PKD1L1	G/C	0.500	0.382	31/72/31	95/287/243	1.62	5.25E-04
	rs7385804	7	100,235,970	TFR2	C/A	0.456	0.378	27/70/39	91/290/244	1.42	1.35E-02
	rs2418135	9	113,901,309	OR2K2	G/A	0.618	0.480	19/66/51	172/306/147	1.82	2.20E-05
	rs61753080	11	119,005,003	HINFP	A/G	0.007	0.004	0/2/133	0/5/616	1.85	3.64E-01
	rs11116595	12	85,165,879	SLC6A15	A/G	0.382	0.436	22/60/54	114/317/194	0.79	9.77E-02
	rs17474506	17	38,990,780	TMEM99	G/C	0.085	0.050	3/17/116	3/57/565	1.74	4.10E-02
Michgan											
	bs1_156907115	1	156,907,115	ARHGEF11	G/A	0.422	0.336	29/93/57	56/235/225	1.46	3.35E-03
	bs2_233990575	2	233,990,575	INPP5D	A/G	0.008	0.000	0/3/176	0/0/517		1.69E-02
	rs115393439	2	233,990,592	INPP5D	C/A	0.008	0.001	0/3/176	0/1/516	8.73	5.47E-02
	rs464494	5	76,003,258	IQGAP2	A/G	0.358	0.460	20/88/71	113/250/154	0.66	1.16E-03
	rs61740965	5	81,608,563	ATP6AP1L	G/A	0.017	0.004	0/6/173	0/4/513	4.39	2.23E-02
	rs10274334	7	47,925,331	PKD1L1	G/C	0.441	0.431	38/82/59	89/268/160	1.02	8.65E-01
	rs7385804	7	100,235,970	TFR2	C/A	0.455	0.374	42/79/58	85/217/215	1.34	1.40E-02
	rs2418135	9	113,901,309	OR2K2	G/A	0.500	0.480	45/89/45	116/263/137	1.09	4.71E-01
	rs61753080	11	119,005,003	HINFP	A/G	0.020	0.005	0/7/170	0/5/510	4.14	1.58E-02
	rs11116595	12	85,165,879	SLC6A15	A/G	0.377	0.485	31/73/75	119/263/135	0.64	4.55E-04
	rs17474506	17	38,990,780	TMEM99	G/C	0.089	0.052	1/30/148	1/52/463	1.78	1.53E-02
CAPS											
	bs1_156907115	1	156,907,115	ARHGEF11	G/A	0.403	0.368	23/74/52	39/140/117	1.17	2.96E-01
	bs2_233990575	2	233,990,575	INPP5D	A/G	0.010	0.000	0/3/146	0/0/296		3.73E-02
	rs115393439	2	233,990,592	INPP5D	C/A	0.023	0.007	0/7/142	0/4/292	3.54	4.96E-02
	rs464494	5	76,003,258	IQGAP2	A/G	0.366	0.411	19/71/59	47/149/100	0.84	2.50E-01
	rs61740965	5	81,608,563	ATP6AP1L	G/A	0.003	0.002	0/1/148	0/1/295	1.99	0.99
	rs10274334	7	47,925,331	PKD1L1	G/C	0.534	0.444	45/69/35	60/143/93	1.46	8.54E-03
	rs7385804	7	100,235,970	TFR2	C/A	0.453	0.439	26/83/40	50/160/86	1.13	4.32E-01
	rs2418135	9	113,901,309	OR2K2	G/A	0.540	0.453	43/75/31	65/138/93	1.39	2.05E-02
	rs61753080	11	119,005,003	HINFP	A/G	0.023	0.007	1/5/143	0/4/291	3.52	4.99E-02
	rs11116595	12	85,165,879	SLC6A15	A/G	0.389	0.476	25/66/58	63/155/77	0.72	2.81E-02
	rs17474506	17	38,990,780	TMEM99	G/C	0.074	0.052	1/20/128	1/29/266	1.44	2.30E-01
Pooled											
	bs1_156907115	1	156,907,115	ARHGEF11	G/A	0.412	0.347	69/244/151	175/646/615	1.33	3.68E-04

bs2_233990575	2	233,990,575	INPP5D	A/G	0.009	0.000	0/8/456	0/0/1437		1.23E-05
rs115393439	2	233,990,592	INPP5D	C/A	0.016	0.003	0/15/449	0/9/1428	5.23	7.86E-05
rs464494	5	76,003,258	IQGAP2	A/G	0.365	0.449	61/217/186	280/730/427	0.72	5.01E-05
rs61740965	5	81,608,563	ATP6AP1L	G/A	0.014	0.003	0/13/451	0/10/1427	4.07	9.41E-04
rs10274334	7	47,925,331	PKD1L1	G/C	0.488	0.412	114/223/125	244/697/496	1.32	2.83E-04
rs7385804	7	100,235,970	TFR2	C/A	0.455	0.389	95/232/137	226/666/545	1.29	9.36E-04
rs2418135	9	113,901,309	OR2K2	G/A	0.547	0.474	139/230/95	328/706/402	1.37	4.81E-05
rs61753080	11	119,005,003	HINFP	A/G	0.017	0.005	1/14/446	0/14/1416	3.59	7.98E-04
rs11116595	12	85,165,879	SLC6A15	A/G	0.383	0.462	78/199/187	296/735/406	0.72	2.92E-05
rs17474506	17	38,990,780	TMEM99	G/C	0.083	0.052	5/67/392	5/138/1293	1.67	7.34E-04

Among those 11 variants, a rare but recurrent missense genetic variant bs2_233990575 in the *INPP5D* region had consistent effect on PCa aggressiveness among all of the 3 populations at a liberal P-value of 0.05 (Table 2). It is located at 233,990,575 bp of chromosome 2, the bs2_233990575 rare allele 'A' appeared only in aggressive PCa subjects, with a minor allele frequency of 0.008, 0.008 and 0.010 in JHH, MI and CAPS populations, respectively. On contrast, this rare allele was not observed among the 625, 517 and 296 non-aggressive PCa subjects in JHH, MI and CAPS populations, respectively (Table 2).

Therefore, we examined the other genetic variants located in the *INPP5D* gene region. We found another rare missense variant rs115393439, which was 17 bp upstream of bs2_233990575 and significantly associated with PCa aggressiveness in the JHH and CAPS populations, with P of 0.012 and 0.050, respectively (Table 2). In addition, rs115393439 was associated with PCa aggressiveness with a marginal P = 0.055 in the MI population (Table 2). Similar with bs2_233990575, the minor allele frequency of rs115393139 was higher in aggressive PCa subjects than in non-aggressive PCa cases, resulting in an odds ratio (OR) of 5.83, 8.73 and 3.54 in JHH, MI and CAPS, respectively (Table 2).

The 11 variants significantly associated with PCa aggressiveness in stage 1 were selected for confirmation study in additional subjects of European descendant from CAPS, including 1,064 subjects with aggressive PCa and 1,291 subjects with non-aggressive PCa. One variant, rs10274334 on chromosome 7 failed the probe design. The remaining 10 variants were successfully genotyped and statistical analysis was performed to test the association between those 10 variants with PCa aggressiveness. We found that the rare variants rs115393439 in *INPP5D* gene and rs61753080 in *HINFP* gene showed significant association ($P<0.05$, Table 3), with an OR of 1.96 and 1.72, respectively. The variant rs115393439 leads to the amino acid substitution from Threonine (Thr) to Proline (Pro); while rs61756080 results to the amino acid substitution from Glycine (Gly) to Glutamic Acid (Glu).

Table 3. Associations between variants and prostate cancer aggressiveness in additional 1,064 aggressive and 1,291 indolent PCa cases from CAPS population

SNP	CHR	BP	Gene	A1/A2	MAF Agg	MAF Indolent	Genotype Agg	Genotype Indolent	OR	P-value
bs1_156907115	1	156,907,115	ARHGEF11	G/A	0.352	0.344	141/467/456	154/579/558	1.04	0.49
bs2_233990575	2	233,990,575	INPP5D	A/G	0.002	0.002	0/4/1060	0/6/1285	0.81	0.98
rs115393439	2	233,990,592	INPP5D	C/A	0.015	0.008	0/32/1032	0/20/1271	1.96	**0.02**
rs464494	5	76,003,258	IQGAP2	A/G	0.409	0.409	163/544/357	217/623/451	1.01	0.85
rs61740965	5	81,608,563	ATP6AP1L	G/A	0.006	0.005	0/12/1052	0/13/1278	1.12	0.84
rs7385804	7	100,235,970	TFR2	C/A	0.429	0.438	173/566/324	260/610/419	0.95	0.44
rs2418135	9	113,901,309	OR2K2	G/A	0.477	0.473	236/542/286	296/629/365	1.00	0.97
rs61753080	11	119,005,003	HINFP	A/G	0.018	0.010	0/38/1026	1/25/1265	1.72	**0.03**
rs11116595	12	85,165,879	SLC6A15	A/G	0.461	0.454	220/540/304	268/636/387	1.02	0.71
rs17474506	17	38,990,780	TMEM99	G/C	0.056	0.057	4/112/948	7/133/1151	0.99	0.95

Gene-based association analysis.

In addition to single variant analysis, we performed the gene-based association analysis in each population using SKAT. All polymorphic variants that passed quality control were included in the analysis. We found there were 4 genes, 3 genes and 1 gene significantly associated with PCa aggressiveness (p-value < 1E-4) in the JHH, MI and CAPS population, respectively (Supplementary Table 3). In the JHH population, the gene *CREB3L1* (cAMP Responsive Element Binding Protein 3-like 1), KLF13 (Kruppel-like Factor 13), *ROBO4* (Roundabout, Axon Guidance Receptor, Homolog 4), and *ZCCHC6* (Zinc Finger, CCHC Domain Containing 6) presented significant association; in the Michigan population, the significant genes were TEK (Tyrosine Kinase, Endothelial), *CDH2* (Cadherin 2), and *BEST2* (Bestrophin 2); while in the CAPS population, the gene showed significant association with PCa aggressiveness was actually a pseudogene LOC100128542. We then explored if there were genes contributing to PCa aggressiveness in at least two populations, setting a p-value threshold of 1E-3. We found 30 genes significantly associated with PCa aggressiveness in the JHH population (p-value < 1E-3), 22 genes in the MI population, and 70 genes in the CAPS population (Table 4). However, none of these genes that were significant at a P-value of 1E-03 were shared in more than 1 population.

Table 4. Gene-based analysis in JHH, MI and CAPS using SKAT.

Population	SetID	P.value	N.Marker.All	N.Marker.Test
JHH				
	CREB3L1	2.55E-06	9	9
	KLF13	1.43E-05	1	1
	ROBO4	2.63E-05	9	9
	ZCCHC6	4.17E-05	7	7
	RNF208	1.01E-04	2	2
	LOC152742	1.24E-04	2	2
	TRIM17	1.72E-04	3	3
	L3MBTL2	1.78E-04	4	4
	SNX10	2.01E-04	2	2
	CXorf68	2.01E-04	1	1
	ZSCAN23	2.01E-04	1	1
	F8	2.01E-04	2	2
	FAM45A	2.28E-04	1	1
	KRTAP22-1	2.63E-04	2	2
	RSG1	2.88E-04	3	3
	TMEM177	3.11E-04	7	7
	CDH6	3.32E-04	4	4
	SPAG7	3.40E-04	2	2
	RAB26	3.48E-04	3	3
	IL16	3.70E-04	12	12
	ZNF829	4.24E-04	3	3
	EXOC3L2	4.32E-04	2	2
	RIMS3	5.03E-04	3	3
	MIR4697	5.60E-04	1	1
	ARHGEF10	6.36E-04	12	12
	C9orf135	6.41E-04	8	8
	MLXIPL	6.65E-04	6	6
	PNMA2	6.73E-04	3	3
	CCL16	9.79E-04	1	1
	AHNAK	9.98E-04	32	32
Michigan				
	TEK	1.47E-05	9	9
	CDH2	5.24E-05	14	14
	BEST2	7.97E-05	4	4
	LOC100130581	1.45E-04	1	1
	OR11L1	2.38E-04	8	8
	S100PBP	2.59E-04	4	4
	LOC643339	4.25E-04	2	2
	INMT	5.07E-04	11	11
	LOC148145	5.51E-04	1	1
	NRIP3	5.80E-04	3	3
	PPARGC1B	6.33E-04	13	13
	TIMM44	7.05E-04	8	8
	LOC401164	7.18E-04	3	3
	RFC1	7.47E-04	2	2
	SLC16A5	7.47E-04	2	2
	SRSF1	7.71E-04	1	1
	MORN3	7.79E-04	4	4

	CDH3	7.79E-04	10	10
	DDHD1	8.71E-04	2	2
	TMEM106C	9.02E-04	5	5
	KLK15	9.52E-04	4	4
	LINC00284	9.80E-04	1	1
CAPS				
	LOC100128542	6.03E-05	2	2
	HS3ST2	2.04E-04	2	2
	PCDH8	2.47E-04	4	4
	MRPL9	2.55E-04	7	7
	PIGA	2.85E-04	1	1
	CCNI2	3.29E-04	1	1
	MGC45800	3.40E-04	3	3
	ZNF624	3.79E-04	2	2
	MAD2L1BP	5.16E-04	1	1
	KIAA1462	5.41E-04	9	9
	BTN2A1	5.95E-04	4	4
	LOC645206	9.17E-04	2	2
	WDR72	9.19E-04	13	13

African American population

Association analysis for single variant in the discovery stage.

We investigated the associations between genetic variants and PCa aggressiveness using ExomeArray in African Americans from JHH study. Although no single variant reached genome-wide significance (P<5×10-7), we found 16 variants associated with PCa aggressiveness with P<1×10-3 (Table 5).

Table 5. Associations between variants and prostate cancer aggressiveness in African American men in JHH population.

SNP	CHR	BP	Gene	A1/A2	MAF Agg	MAF Non-Agg	Genotype Agg	Genotype Non-Agg	OR	P-value
rs663824	1	43,649,508	*WDR65*	A/G	0.440	0.130	5/12/2008	0/7/20	5.28	4.21E-04
rs4147825	1	94,560,938	*ABCA4*	A/G	0.173	0.482	1/7/2018	7/12/2008	0.23	7.37E-04
rs1801274	1	161,479,745	*FCGR2A*	A/G	0.269	0.630	1/12/2013	12/10/2005	0.22	1.94E-04
rs1564348	6	160,578,860	*SLC22A1*	G/A	0.231	0.019	1/10/2015	0/1/26	15.90	8.67E-04
rs4947385	7	51,963,775	*COBL*	G/A	0.269	0.593	1/12/2013	11/10/2006	0.25	7.85E-04
rs28750165	7	107,616,188	*LAMB1*	A/G	0.077	0.365	0/4/22	4/11/2011	0.14	3.94E-04
rs590937	10	43,149,991	*ZNF33B*	A/G	0.269	0.593	2/10/2014	6/20/2001	0.25	7.85E-04
rs7095762	10	115,910,928	*C10orf118*	A/C	0.519	0.185	6/15/2005	0/10/17	4.75	3.10E-04
rs1061159	10	115,922,774	*C10orf118*	A/G	0.500	0.185	6/14/2006	0/10/17	4.40	6.23E-04
rs9664945	10	116,008,497	*VWA2*	A/G	0.500	0.185	6/14/2006	0/10/17	4.40	6.23E-04
rs1908946	12	25,243,115	*LRMP*	G/C	0.080	0.352	0/4/21	4/11/2012	0.16	8.45E-04
rs2306480	15	40,539,373	*PAK6*	A/G	0.654	0.333	13/8/5	2/14/2011	3.78	9.67E-04
rs1197682	15	42,171,483	*SPTBN5*	A/G	0.019	0.278	0/1/25	1/13/2013	0.05	2.02E-04
rs890499	15	42,179,424	*SPTBN5*	A/G	0.019	0.259	0/1/25	0/14/13	0.06	3.93E-04
rs28418770	19	57,931,425	*ZNF17*	G/A	0.077	0.352	0/4/22	4/11/2012	0.15	5.97E-04
rs6136489	20	1,923,734	*SIRPA*	A/C	0.442	0.111	5/13/2008	0/6/21	6.35	1.31E-04

Association Analysis in Additional African Americans in confirmation stag

The 16 variants significantly associated in stage 3 were selected for confirmation in an additional 772 African Americans from the JHH study, including 388 subjects with aggressive PCa and 384 subjects with non-aggressive PCa. In addition to the 16 variants associated with PCa aggressiveness in AAs, the significant variants confirmed in Caucasians in stage 2, rs115393439 in *INPP5D* and rs61758030 in *HINFP*, were also selected. Moreover, we also evaluated the association between an additional 7 variants in *INPP5D* and 3 variants in *HINFP* that were polymorphic in AAs, and PCa aggressiveness. Among the 28 variants selected, 2 of them failed in probe design, and 1 variant, rs11539439 was not polymorphic in African Americans. Therefore, 26 SNPs remained for statistical analysis. We found that 2 variants, rs75905572 and rs183287568 in HINFP were significantly associated with PCa aggressiveness (P<0.05). The allele "G" of rs75905572, was present less frequent in the aggressive PCa (4.2%), compared with indolent PCa (6.7%), with a P-value of 0.046. Men who carry the "G" allele had 0.61 fold decreased risk for aggressive PCa, compared with men carrying the "C" allele (OR = 0.61, P= 0.046). The allele "C" of rs183287568, was present less frequent in the aggressive PCa (1.1 %), compared with indolent PCa (0.1%), with a P-value of 0.039. Men who carry the "C" allele had 7.74 fold increased risk for aggressive PCa, compared with men carrying the "T" allele (OR = 7.74< P =0.039)

Table 6. Associations between variants and prostate cancer aggressiveness in stage 4.

SNP	CHR	BP	Gene	A1/A2	MAF Agg	MAF Indolent	Genotype Agg	Genotype Indolent	OR	P-value
rs663824	1	43,649,508	*WDR65*	A/G	0.218	0.238	23/110/224	26/111/205	0.89	0.41
rs4147825	1	94,560,938	*ABCA4*	T/C	0.339	0.346	42/159/156	42/153/146	0.97	0.78
rs1801274	1	161,479,745	*FCGR2A*	A/G	0.455	0.462	70/186/101	69/178/95	0.98	0.87
rs114254639	2	233,989,671	*INPP5D*	G/C	0.004	0.004	0/3/354	0/3/339	0.96	0.99
rs145592503	2	233,989,802	*INPP5D*	T/C	0.011	0.012	0/8/348	0/8/334	0.96	0.99
rs148905765	2	233,989,933	*INPP5D*	A/G	0.020	0.010	0/14/343	0/7/335	1.93	0.18
rs142612494	2	233,990,091	*INPP5D*	T/C	0.006	0.001	0/4/353	0/1/341	3.85	0.38
rs112464218	2	233,990,295	*INPP5D*	A/G	0.025	0.020	1/16/340	1/12/327	1.23	0.59
rs1564348	6	160,578,860	*SLC22A1*	C/T	0.113	0.113	7/67/283	13/51/278	1.01	0.99
rs4947385	7	51,963,775	*COBL*	C/T	0.404	0.390	61/167/129	57/153/132	1.06	0.58
rs28750165	7	107,616,188	*LAMB1*	T/C	0.151	0.168	7/94/256	9/97/236	0.88	0.42
rs590937	10	43,149,991	*ZNF33B*	T/C	0.446	0.428	72/175/108	68/157/117	1.09	0.45
rs7095762	10	115,910,928	*C10orf118*	A/C	0.295	0.294	39/133/185	24/153/165	1.01	0.95
rs9664945	10	116,008,497	*VWA2*	A/G	0.303	0.284	37/143/176	23/148/166	1.08	0.51
rs61753080	11	119,005,003	*HINFP*	A/G	0.003	0.004	0/2/355	0/3/339	0.64	0.68
rs114318772	11	119,005,057	*HINFP*	A/G	0.006	0.003	0/4/347	0/2/338	1.94	0.69
rs75905572	11	119,005,900	*HINFP*	G/C	0.042	0.067	1/28/327	1/44/297	0.61	**0.046**

rs183287568	11	119,006,298	*HINFP*	C/T	0.011	0.001	0/8/349	0/1/341	7.74	**0.039**
rs1908946	12	25,243,115	*LRMP*	C/G	0.244	0.254	19/137/201	24/126/192	0.95	0.71
rs2306480	15	40,539,373	*PAK6*	A/G	0.426	0.444	67/171/118	67/170/104	0.93	0.52
rs890499	15	42,179,424	*SPTBN5*	A/G	0.175	0.189	11/103/242	19/91/230	0.91	0.53
rs28418770	19	57,931,425	*ZNF17*	G/A	0.197	0.225	16/109/231	19/116/206	0.85	0.21
rs6136489	20	1,923,734	*SIRPA*	T/G	0.320	0.346	37/155/165	34/169/139	0.89	0.31

Gene-based analysis

We also performed gene-based analysis using the SKAT approach in the AA population. The top genes with P-values < 1E-03 are presented in Table 7. We conducted the SKAT analysis based on all variants. A total of 33 genes sets were identified (Table 7). The top gene sets associated with aggressive PCa were *JOSD1, C10orf118 and PHEX,* with P-values that ranged from 3.95E-04 to 4.82E-04.

Table 7. Top signficant genes associated with aggressive PCa using SKAT approach in AAs from JHH population (based on all variants)

SetID	P.value	N.Marker.All	N.Marker.Test
JOSD1	0.000395	2	2
C10orf118	0.000463	2	2
PHEX	0.000482	3	3
FOXP2	0.001433	2	2
MIR663A	0.002451	2	2
MBTPS2	0.002472	1	1
UCN3	0.0026	1	1
SOX14	0.00356	4	4
OTOS	0.003864	1	1
LOC100133050	0.003864	4	4
POM121L4P	0.003901	1	1
LOC339593	0.004047	4	4
CRY1	0.004529	2	2
HTR2C	0.004545	1	1
C9orf4	0.004659	1	1
ELF4	0.005329	1	1
KLHL14	0.006274	1	1
GPAM	0.006491	7	7
PDE3A	0.006595	5	5
C20orf203	0.00676	1	1
CELF2	0.007475	1	1
UBL5	0.007773	1	1
C1orf114	0.007844	1	1
MIR1252	0.008487	1	1

MGC12916	0.00865	3	3
NBPF3	0.008682	4	4
EMB	0.008763	1	1
LOC283624	0.008908	1	1
TMED11P	0.009345	1	1
C14orf133	0.009361	1	1
MIR4272	0.009538	2	2
EPGN	0.009573	1	1

Discussion

To our knowledge, our study represents one of the first comprehensive studies to identify rare variants that are associated with aggressive PCa in both EAs and AAs. Our data generated from the entire funding period identified novel rare variants that were associated with aggressive PCa. In summary, using a multi-stage study design, we identified two novel rare variants associated with PCa aggressiveness in Caucasians, including 1 in the *INPP5D* gene and 1 in the *HINFP* gene. We also discovered two additional novel rare variants in the *HINFP* gene that were associated with PCa aggressiveness in African Americans. More importantly, those rare variants are located in the coding region of the genes, leading to amino acids changes. The replication of these findings in additional populations indicated that the variants identified may represent truly associated genes with PCa aggressiveness.

We selected the Illumina Human Exome BeadChip (ExomeArray) as our genotyping platform to study rare variants. The ExomeArray chip represents the newest gene chip that delivers unparalleled coverage of putative functional exonic variants. The relatively cheaper cost of this platform makes it possible to study larger sample sizes. The Exome Beachip is comprised of >240,000 markers, including >200,000 nonsynonymous SNPs, nonsense mutations, SNPs in splice sites and promoter regions, as well as thousands of GWAS tag markers. Nearly 90% of the SNPs on the exome arrays are rare, with a MAF<5%. In addition, the markers on the Illumina Human Exome BeadChips are selected from over 12,000 individual exome and whole-genome sequences, representing diverse populations, including those of European and African descent. Therefore, it is more efficient and economical to use exome arrays to identify rare variants associated with aggressive PCa, compared with whole genome sequencing.

The gene Inositol Polyphosphate-5-Phosphatase 1 (*INPP5D*) is a member of INPP5 family. It encodes a Phosphatidylinositol (PtdIns) phosphatase that specifically hydrolyzes the 5-phosphate of phosphatidylinositol (3,4,5)-triphosphate (PtdIns (3,4,5) P3) to produce PtdIns(3,4)P2, and therefore negatively regulating the PI3K (phosphoinositide 3-kinase) pathways. Acting as an inhibitor of the PI3K pathway, INPP5D is considered as a tumor suppressor in acute myeloid leukemia, Hodgkin's lymphoma, and acute lymphoblastic leukemia (Luo et al. 2004; Metzner et al. 2009; Tiacci et al. 2012). Besides, INPP5D has been identified as the target of the cellular tumor antigen p53 in human breast cancer adenocarcinoma MCF7 cells and testicular germ cell tumor-derived human embryonal carcinoma cells (Kerley-Hamilton et al. 2005; Lion et al. 2013). Although the role INPP5D plays in prostate tumor cells has not been

established, it is possible that INPP5D contributed to prostate cancer progression through PI3K or p53 pathway.

The gene Histone H4 Transcription Factor (*HINFP*) is heavily involved in cell cycle progression. It encodes a key transcription factor of histone H4 genes, which play a central role in genome replication and stability (Marzluff et al. 2002; Mitra et al. 2003). HINFP is essential for E2F-independent activation of the histone H4 gene family (Mitra et al. 2003). Responding to the Cyclin E / Cyclin-dependent Kinase 2 cell cycle signaling, HINFP binds to p220, and thus regulates histone H4 gene transcription at the G1/S phase transition (Miele et al. 2005; Mitra et al. 2003). As Cyclin E / CDK2-dependent mechanisms contribute to prostate cancer cell proliferation (Flores et al. 2010), the resulting pathological consequences may be mediated through HINFP. In addition, HINFP interacts directly with methyl-CpG-binding protein-2 (MBD2) (Sekimata et al. 2001), which is repressed in prostate cancer (Patra et al. 2002). Since MBD2 is heavily involved in DNA methylation mediated transcription repression (Sekimata et al. 2001), HINFP may contribute to prostate cancer progression in this manner. In addition, HINFP directly regulated other cell cycle and cancer related genes, including ATM, PRKDC and CKS2 (Medina et al. 2007), which provide potential mechanism through which HINFP contributes to prostate cancer progression.

In addition to the single variant analysis, we also performed gene-based approach to identify genes associated with PCa aggressiveness. The gene-based approach (SKAT) we adopted is a novel statistical approach. SKAT is a supervised and flexible regression method to test for association between rare variants in a gene or genetic region and a continuous or dichotomous trait. Compared to other methods of estimating the joint effect of a subset of SNPs, SKAT is able to deal with variants that have different direction and magnitude of effects, and allows for covariate adjustment (Wu 2011). In addition, SKAT can also avoid arbitrary selection of threshold in burden test. Moreover, SKAT is computationally efficient, compared to a permutation test, making it feasible to analyze the large dataset in our study. Interestingly, several of the top targets identified by SKAT analysis (*CREB3L1* and *KLF13*) encode transcription factors.

Besides all the above findings, we have also carefully calculated the study power based on our modified study design. We have >80% power to detect an OR of 1.7 (2.8) for variants with a MAF of 0.05 (0.01), at an alpha level of 1E-05 (2-sided). Therefore, we have sufficient power to identify novel rare mutations with relatively large effect based on our proposed sample size. We also considered several procedures to control for multiple test correction and SNP selection to be confirmed in additional independent samples. The Bonferroni corrected P-values are 2E-7 (0.05/200,000 variants) and 2E-6 (0.05/20,000 genes), for single variant analysis and gene-based analysis, respectively. However, not all the tests for single variants are independent due to linkage disequilibrium (LD) structure among variants. In addition, previous studies also showed that the true associations do not necessarily reach the stringent Bonferroni corrected P-value cutoffs. Therefore, to balance study power and false positives, rare variants in Aim 1 that meet either of the following criteria with less stringent P-value cutoffs will be selected for replication: 1) variants reach a p-value of 1E-3 in single variant analysis; 2) variants in genes which reach a p-value of 1E-3 in gene-based analysis by SKAT. The adoption of the two-stage study design will further help to remove false positives.

In conclusion, we have identified several novel rare variants and genes that are associated with aggressive PCa in Caucasians and African American men. The newly identified variants can provide more insight into the etiology of aggressive PCa and provide potential effective targets for therapy of aggressive PCa.

KEY RESEARCH ACCOMPLISHMENTS

1) Completed IRB and other logistical issues
2) Performed single rare variant analysis, bioinformatics analysis, and gene-based analysis (SKAT) to identify rare variants that have strong effects on aggressive PCa risk in exome-array data among a total of 1,919 PCa cases, including 470 aggressive PCa cases and 1,449 indolent PCa cases.
3) Performed replication study in additional 1,421 aggressive PCa cases and 1,633 indolent PCa cases to confirm the variants that implicated in the discovery stage.
4) Successfully identified two novel rare variants associated with PCa aggressiveness in Caucasians, including one (rs115393139) in *INPP5D* geneand one(rs61753080) in *HINFP* gene. OR of rs11593139 is 5.83, 8.73 and 3.54 in JHH, Michigan and CAPS, respectively. OR of rs61753080 is 1.85, 3.52 and 4.14 in JHH, Michigan and CAPS, respectively.
5) Successfully identified two novel rare variants, rs75905572 and rs183287568, in the *HINFP* gene that were associated with PCa aggressiveness in African American men, with OR of 0.61 and 7.74, respectively.

REPORTABLE OUTCOMES

1) Top variants and genes in the genome that are significantly associated with aggressive PCa in EAs (Table 2 - Table 4)
2) Top variants and genes in the genome that are significantly associated with aggressive PCa in AAs (Table 5 - Table 7)

CONCLUSION

1) We have made great progress in achieving the goals described in the approved Statement of Work.
2) We have identified and confirmed several novel rare variant in the *INPP5D* gene and *HINFP* gene that are associated with aggressive PCa in both Caucasians and African American men.
3) The newly identified variants can provide more insight into the etiology of aggressive PCa and provide potential effective targets for therapy of aggressive PCa.

REFERENCES

Adzhubei I.A., Schmidt S., Peshkin L., et al. A method and server for predicting damaging missense mutations. Nat Methods. 7, 248-249 (2010).

Akbari MR, Trachtenberg J, Lee J, Tam S, Bristow R, Loblaw A, Narod SA, Nam RK. Association Between Germline HOXB13 G84E Mutation and Risk of Prostate Cancer. J Natl Cancer Inst. 2012 Aug 1;104(16):1260-2. Epub 2012 Jul 9.

Castro E. G.C.L., Olmos D., et al. Correlation of germ-line BRCA2 mutations with aggressive prostate cancer and outcome. ASCO Meet Abstr. 29, 1517 (2011).

Dunant NM, Wisniewski D, Strife A, Clarkson B, Resh MD (2000) The phosphatidylinositol polyphosphate 5-phosphatase SHIP1 associates with the dok1 phosphoprotein in bcr-Abl transformed cells. Cell Signal 12: 317-26.

Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, Wiley KE, Isaacs SD, Johng D, Wang Y, Bizon C, Yan G, Gielzak M, Partin AW, Shanmugam V, Izatt T, Sinari S, Craig DW, Zheng SL, Walsh PC, Montie JE, Xu J, Carpten JD, Isaacs WB, Cooney KA. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med. 2012 Jan 12;366(2):141-9.

Falk K, Rötzschke O. The final cut: how ERAP1 trims MHC ligands to size. Nat Immunol. 2002 Dec;3(12):1121-2.

Flores O, Wang Z, Knudsen KE, Burnstein KL (2010) Nuclear targeting of cyclin-dependent kinase 2 reveals essential roles of cyclin-dependent kinase 2 localization and cyclin E in vitamin D-mediated growth inhibition. Endocrinology 151: 896-908. doi: 10.1210/en.2009-1116

Fredericks WJ, McGarvey T, Wang H, Lal P, Puthiyaveettil R, Tomaszewski J, Sepulveda J, Labelle E, Weiss JS, Nickerson ML, Kruth HS, Brandt W, Wessjohann LA, Malkowicz SB. The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression. DNA Cell Biol. 2011 Nov;30(11):851-64. Epub 2011 Jul 8.

Gallagher D.J., Gaudet M.M., Pal P., et al. Germline BRCA mutations denote a clinicopathologic subset of prostate cancer. Clin Cancer Res. 16, 2115-2121 (2010).

Hammer GE, Gonzalez F, Champsaur M, Cado D, Shastri N. The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules. Nat Immunol. 2006 Jan;7(1):103-12. Epub 2005 Nov 20.

Heemels MT, Ploegh H. Generation, translocation, and presentation of MHC class I-restricted peptides.
Annu Rev Biochem. 1995;64:463-91. Review.

Henson BJ, Gollin SM. Overexpression of KLF13 and FGFR3 in oral cancer cells. Cytogenet Genome Res. 2010 Jun;128(4):192-8. doi: 10.1159/000308303. Epub 2010 Jun 2.

Hinks A, Martin P, Flynn E, Eyre S, Packham J; Childhood Arthritis Prospective Study-CAPS; BSPAR Study Group, Barton A, Worthington J, Thomson W. Subtype specific genetic associations for juvenile idiopathic arthritis: ERAP1 with the enthesitis related arthritis subtype and IL23R with juvenile psoriatic arthritis. Arthritis Res Ther. 2011 Jan 31;13(1):R12. doi: 10.1186/ar3235.

Kanazawa A, Kawamura Y, Sekine A, Iida A, Tsunoda T, Kashiwagi A, Tanaka Y, Babazono T, Matsuda M, Kawai K, Iiizumi T, Fujioka T, Imanishi M, Kaku K, Iwamoto Y, Kawamori R, Kikkawa R, Nakamura Y, Maeda S. Single nucleotide polymorphisms in the gene encoding Krüppel-like factor 7 are associated with type 2 diabetes. Diabetologia. 2005 Jul;48(7):1315-22. Epub 2005 Jun 4.

Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, Grönberg H, Wiklund F. A Population-based Assessment of Germline HOXB13 G84E Mutation and Prostate Cancer Risk. Eur Urol. 2012 Jul 20. [Epub ahead of print]

Kerley-Hamilton JS, Pike AM, Li N, DiRenzo J, Spinella MJ (2005) A p53-dominant transcriptional response to cisplatin in testicular germ cell tumor-derived human embryonal carcinoma. Oncogene 24: 6090-100. doi: 10.1038/sj.onc.1208755

Lion M, Bisio A, Tebaldi T, De Sanctis V, Menendez D, Resnick MA, Ciribilli Y, Inga A (2013) Interaction between p53 and estradiol pathways in transcriptional responses to chemotherapeutics. Cell Cycle 12: 1211-24. doi: 10.4161/cc.24309

Luo JM, Liu ZL, Hao HL, Wang FX, Dong ZR, Ohno R (2004) Mutation analysis of SHIP gene in acute leukemia. Zhongguo Shi Yan Xue Ye Xue Za Zhi 12: 420-6.

Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ (2002) The human and mouse replication-dependent histone genes. Genomics 80: 487-98.

Medina R, van der Deen M, Miele-Chamberland A, Xie RL, van Wijnen AJ, Stein JL, Stein GS (2007) The HiNF-P/p220NPAT cell cycle signaling pathway controls nonhistone target genes. Cancer Res 67: 10334-42. doi: 10.1158/0008-5472.CAN-07-1560

Mehta AM, Jordanova ES, Corver WE, van Wezel T, Uh HW, Kenter GG, Jan Fleuren G. Single nucleotide polymorphisms in antigen processing machinery component ERAP1 significantly associate with clinical outcome in cervical carcinoma. Genes Chromosomes Cancer. 2009 May;48(5):410-8.

Metzner A, Precht C, Fehse B, Fiedler W, Stocking C, Gunther A, Mayr GW, Jucker M (2009) Reduced proliferation of CD34(+) cells from patients with acute myeloid

leukemia after gene transfer of INPP5D. Gene Ther 16: 570-3. doi: 10.1038/gt.2008.184

Miele A, Braastad CD, Holmes WF, Mitra P, Medina R, Xie R, Zaidi SK, Ye X, Wei Y, Harper JW, van Wijnen AJ, Stein JL, Stein GS (2005) HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition. Mol Cell Biol 25: 6140-53. doi: 10.1128/MCB.25.14.6140-6153.2005

Mitra P, Xie RL, Medina R, Hovhannisyan H, Zaidi SK, Wei Y, Harper JW, Stein JL, van Wijnen AJ, Stein GS (2003) Identification of HiNF-P, a key activator of cell cycle-controlled histone H4 genes at the onset of S phase. Mol Cell Biol 23: 8110-23.

Patra SK, Patra A, Zhao H, Dahiya R (2002) DNA methyltransferase and demethylase in human prostate cancer. Mol Carcinog 33: 163-71.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006 Aug;38(8):904-9. Epub 2006 Jul 23.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC. PLINK: a toolset for whole-genome association and population-based
linkage analysis. American Journal of Human Genetics. 2007, 81.

Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, Greer F, Schomburg L, Fruci D, Niedermann G, van Endert PM. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. Nat Immunol. 2005 Jul;6(7):689-97. Epub 2005 May 22.

Szczypiorska M, Sánchez A, Bartolomé N, Arteta D, Sanz J, Brito E, Fernández P, Collantes E, Martínez A, Tejedor D, Artieda M, Mulero J. ERAP1 polymorphisms and haplotypes are associated with ankylosing spondylitis susceptibility and functional severity in a Spanish population. Rheumatology (Oxford). 2011 Nov;50(11):1969-75. doi: 10.1093/rheumatology/ker229. Epub 2011 Aug 24.

Sekimata M, Takahashi A, Murakami-Sekimata A, Homma Y (2001) Involvement of a novel zinc finger protein, MIZF, in transcriptional repression by interacting with a methyl-CpG-binding protein, MBD2. J Biol Chem 276: 42632-8. doi: 10.1074/jbc.M107048200

Scohy S, Gabant P, Van Reeth T, Hertveldt V, Drèze PL, Van Vooren P, Rivière M, Szpirer J, Szpirer C. Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF transcription factor family.
Genomics. 2000 Nov 15;70(1):93-101.

Tiacci E, Doring C, Brune V, van Noesel CJ, Klapper W, Mechtersheimer G, Falini B, Kuppers R, Hansmann ML (2012) Analyzing primary Hodgkin and Reed-Sternberg cells to capture the molecular and cellular pathogenesis of classical Hodgkin lymphoma. Blood 120: 4609-20. doi: 10.1182/blood-2012-05-428896

Thorne H., Willems A.J., Niedermayr E., et al. Decreased prostate cancer-specific survival of men with BRCA2 mutations from multiple breast cancer families. Cancer Prev Res (Phila). 4, 1002-1010 (2011).

Trembath RC Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2, Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, Barton A, Band G, Bellenguez C, Bergboer JG, Blackwell JM, Bramon E, Bumpstead SJ, Casas JP, Cork MJ, Corvin A, Deloukas P, Dilthey A, Duncanson A, Edkins S, Estivill X, Fitzgerald O, Freeman C, Giardina E, Gray E, Hofer A, Hüffmeier U, Hunt SE, Irvine AD, Jankowski J, Kirby B, Langford C, Lascorz J, Leman J, Leslie S, Mallbris L, Markus HS, Mathew CG, McLean WH, McManus R, Mössner R, Moutsianas L, Naluai AT, Nestle FO, Novelli G, Onoufriadis A, Palmer CN, Perricone C, Pirinen M, Plomin R, Potter SC, Pujol RM, Rautanen A, Riveira-Munoz E, Ryan AW, Salmhofer W, Samuelsson L, Sawcer SJ, Schalkwijk J, Smith CH, Ståhle M, Su Z, Tazi-Ahnini R, Traupe H, Viswanathan AC, Warren RB, Weger W, Wolk K, Wood N, Worthington J, Young HS, Zeeuwen PL, Hayday A, Burden AD, Griffiths CE, Kere J, Reis A, McVean G, Evans DM, Brown MA, Barker JN, Peltonen L, Donnelly P, Trembath RC. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1.Nat Genet. 2010 Nov;42(11):985-90. doi: 10.1038/ng.694. Epub 2010 Oct 17.

Tryggvadottir L., Vidarsdottir L., Thorgeirsson T., et al. Prostate cancer progression and survival in BRCA2 mutation carriers. J Natl Cancer Inst. 99, 929-935 (2007).

Wu M.C., Lee S., Cai T., et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 89, 82-93 (2011).

Wan X, Yang C, Yang Q, Xue H, Tang NL, Yu W.Predictive rule inference for epistatic interaction detection in genome-wide association studies. Bioinformatics. 2010 Jan 1;26(1):30-7. Epub 2009 Oct 30.

Yan J, Parekh VV, Mendez-Fernandez Y, Olivares-Villagómez D, Dragovic S, Hill T, Roopenian DC, Joyce S, Van Kaer L. In vivo role of ER-associated peptidase activity in tailoring peptides for presentation by MHC class Ia and class Ib molecules. J Exp Med. 2006 Mar 20;203(3):647-59. Epub 2006 Feb 27.

Zhang,Y. and Liu,J.S. (2007) Bayesian inference of epistatic interactions in case-control studies. Nat. Genet., 39,1167–1173.